Properties/Sub-systems

Properties: Evaluation of genes in context.

- comF is found in species that are and are not competent
- In E.coli, it has been shown to be involved in gluconate metabolism
- comF is possibly as a transport permease
- → Without knowing the phenotype of the organism homology alone does not tell you everything

H. flu	E. coli	Similarity
comA	yrfD	33.2
comB	yrfC	59.2
comC	yrfB	49.7
comD	yrfA	44.2
comE	hofQ	68.9
comF	yhgH	69.5
comG	yhgI	91.9
comI	yhiR	85.1

What is a property/sub-system?

An *attribute* of an organism that will go by many names among biologists and computational geeks, but hopefully they are…rigorously defined

- so that the assertion of its absence, presence is meaningful
- can be made either automatically or manually
- can be quantitative
- probably maps to a CV term, but doesn't yet.

Examples

Biological niche:
- animal pathogen, human pathogen, optimum salinity, optimal growth temperature, ptimal pH, oxygen requirement, plant pathogen, temperature environment

Cell surface component:
- capsule, flagella, outer membrane, peptidoglycan(murein) biosynthesis, S-layer, type IV pilus

Cellular growth, organization and division , cell shape, small molecule transport
- count of TRAP transporter clusters, K+-transporting ATPase KdpFABC, Na+-translocating NADH-quinone reductase, NADH dehydrogenase I, phosphate ABC transporter (pstSCAB-phoU), PTS transport system, sulfate/thiosulfate ABC transporter, TRAP-T (tripartite ATP-indep. periplasmic) transporters

Protein transport
- Sec-system protein translocase, Tat (Sec-independent) protein export, type I-IV secretion

Quantitative content
- amino acid abundance, count of DNA molecules, count of predicted proteins, count of tRNAs, DNA dinucleotide thermophily metric RR+YY-RY-YR, DNA GC content, DNA size (megabases), functional gene clustering - property level, protein average length

Selfish genetic elements
- CRISPR region, group I intron, group II intron, inteins

Metabolism
- biosynthesis, catabolism, central intermediary metabolism, energy metabolism, nucleic acid metabolism, protein modification and cofactors, storage polymer systems

Defining Properties

- E.g., com locus
  - Gluconate metabolism v. competency
- Requires evaluation of a *group* of genes
- Do not necessarily map to single genes
- Pathway definitions
  - Typically include yes, no, ORs, maybe, and subsets of gene collections
  - e.g., TCA cycle

GO ←→Sub-systems

- Part of the reason GO assignments are hard

- Note above datatypes (GC content, phylogeny) outside of GO

- No current implentation for making GO assignments

TIGR Model

- "Genome Properties"
- Rules:
  - Subsets w/ encoding of yes, no, or, maybe
- Curatorial and automated components
- Comparison versus other gold standards
- *Capability* to link to GO, not a lot of data
- Displayed on CMR

Comparative Species Analysis

Properties	Species and Taxa showing indicated pattern of property states							
	most γ-prote o-bacte ria	Bacilli, Actino-bacteria, *Bacteroides*, *Chlorobium*, *Haemophilus*	α-, β-proteobacte ria, *Pseudomonas*	*Bifido-bacterium*, *Clostridium*, *Streptococcus*	Archaea, α-, ε-proteobacteria, *Deinococcus*, *B. halodurans, Thermo-anaerobacter, Aquifex*, *Pirellula*, *Streptomyces*, *Thermotoga*, Xanthomonad ales	*Aeropyrum pernix, Pyrococcus abyssi*	`Chlamy dia`	Spirochaetes, *Mycoplasma*, *Rickettsia*, *Wigglesworth ia*, *Pyrococcus horikoshii*
Chorismate biosynthesis	YES	YES	YES	YES	YES	YES	YES	no[a]
Tryptophan biosynthesis	YES	YES	YES	YES	YES	YES	no	no
Phenylalanine biosynthesis	YES	YES	YES	YES	YES	no	no	no
Tyrosine biosynthesis	YES	YES	YES	YES	YES	no	no	no
Menaquinone biosynthesis	YES	YES	no	no	no	no	no	no
Ubiquinone biosynthesis[b]	YES	no	YES	no	some evidence	some evidence	some evidence	no

Curation v automated

- Ian Paulsen: transporters
- Berges Manual
- TIGR/Ross: positive examples

Comparative Species Analysis

Properties	Species and Taxa showing indicated pattern of property states							
	most γ-proteo-bacteria	Bacilli, Actino-bacteria, *Bacteroides, Chlorobium, Haemophilus*	α-, β-proteobacteria, *Pseudomonas*	*Bifido-bacterium, Clostridium, Streptococcus*	Archaea, α-, ε-proteobacteria, *Deinococcus, B. halodurans, Thermo-anaerobacter, Aquifex, Pirellula, Streptomyces, Thermotoga,* Xanthomonadales	*Aeropyrum pernix, Pyrococcus abyssi*	`Chlamydia`	Spirochaetes, *Mycoplasma, Rickettsia, Wigglesworthia, Pyrococcus horikoshii*
Chorismate biosynthesis	YES	YES	YES	YES	YES	YES	YES	no[a]
Tryptophan biosynthesis	YES	YES	YES	YES	YES	YES	no	no
Phenylalanine biosynthesis	YES	YES	YES	YES	YES	no	no	no
Tyrosine biosynthesis	YES	YES	YES	YES	YES	no	no	no
Menaquinone biosynthesis	YES	YES	no	no	no	no	no	no
Ubiquinone biosynthesis[b]	YES	no	YES	no	some evidence	some evidence	some evidence	no